The RTES Project - BTeV, and Beyond

Michael J Haney (m-haney@uiuc.edu), for
The RTES Collaboration

Summary

The Real Time Embedded Systems (RTES) project was created to study the design and implementation of high-performance, heterogeneous, self-aware and fault-adaptive real-time embedded systems. The driving application for this research was the proposed BTeV high energy physics experiment.

At the time of inception for RTES, the Level 1 BTeV trigger was to have required a farm of approximately 2500 DSPs for the pixel trigger algorithm, and an additional (identical) farm of 250 DSPs for the muon trigger algorithm. The Level 2/3 trigger was to have required approximately 2500 conventional Linux computers. There was a concern that developing and maintaining a system of this magnitude and complexity could be prohibitive, and that strong modeling tools (for development) and a reliable/fault-adaptive infrastructure (for runtime) were required.

The RTES project pulled together computer science and electrical engineering expertise from several institutions to address these concerns. The RTES approach incorporates 3 primary elements:

- GME (generic modeling environment) graphical models for the system integration level, message types, run control state machines, user interface definition, custom ARMOR elements (below), as well as metamodels which define the domain-specific graphical languages for each of the above.

- ARMORs (adaptive reconfigurable mobile objects for reliability) which provide intrinsically reliable high level process oversight, with the ability to checkpoint, stop, and restart applications in-place (on the same node), or to relocate the work to an available alternative node. ARMORs insure that protected processes are always running, somewhere, and ARMORs protect themselves as protected processes. ARMORs are modular, and support custom-developed "elements" which can have application-specific fault sensitivities and mitigation behaviors beyond the simple checkpointed stop/restart.

- VLAs (very lightweight agents) which provide low impact, low level detection of fault conditions, and via Elvin communication can collaborate with other VLAs and ARMOR custom elements to effect local, regional, and global responses to fault conditions.

A prototype for the BTeV Level 1 trigger was developed and presented at SuperComputing 2003, running on a small number of DSPs, but capturing the full scope of the command/control, monitoring and fault-adaptive aspects of the RTES solution.

At the time of BTeV's termination early in 2005, the RTES project was within days of completing a significant prototype implementation for providing reliability and fault-adaptive infrastructure to the L2/3 farm. Unlike the L1 prototype (in 2003), which was developed on RTES-specific DSP hardware, the L2/3 prototype was developed for the 84+ nodes of the BTeV Level 2/3 protofarm in service at Fermilab.

This paper will document the conclusion of the RTES project's focus on BTeV, with a brief summary of the L1 prototype results, and salient details from the L2/3 system development.

The paper will also provide an evaluation of the applicability of the RTES concepts to other systems. At the time of writing this summary, RTES has already been (initially) examined for applicability to the data acquisition software of the Dark Energy Survey (DES), and discussions have begun for studies of applicability to the CMS experiment. Other candidate systems are also being considered, and will be examined in this paper.